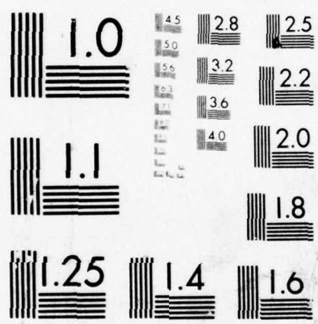
AD-A072 406 SOUTHERN METHODIST UNIV DALLAS TX DEPT OF OPERATIONS--ETC F/G 12/1
ANOTHER VERSATILE FAMILY OF PROBABILITY DISTRIBUTIONS.(U)
DEC 78 B W SCHMEISER, R LAL N00014-77-C-0425
UNCLASSIFIED OREM-78015 NL

| OF |
AD
A072406

END
DATE
FILMED
9-79
DDC



MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963-A

LEVEL II

12 B.S.

AD A072406

DDC FILE COPY

Technical Report OREM 78015

See 1473 in back

DDC
RECEIVED
AUG 8 1979
C

ANOTHER VERSATILE FAMILY OF PROBABILITY DISTRIBUTIONS

Bruce W. Schmeiser

Ram Lal

Department of Operations Research
and
Engineering Management
Southern Methodist University
Dallas, Texas 75275

December 1978

This research was supported by the Office of Naval Research, Code 431, Contract N00014-77-C-0425.

79 08 06 084

ABSTRACT

A new four parameter family of probability distributions is described. Special cases include Bernoulli trials, uniform, power series, exponential, triangular and Laplace (double exponential) distributions. Statistical properties, parameter determination and random variate generation are discussed.

KEY WORDS

Family of Distributions

Simulation

Process Generation

Modeling

I. INTRODUCTION

Considered in this paper is a four parameter family of probability distributions having density function

$$
\begin{aligned}
f(x) &= [1+(x-\lambda_1)/(\lambda_2\lambda_4)]^{(1-\lambda_3)/\lambda_3}/(\lambda_2\lambda_3) && \text{if } \lambda_1-\lambda_2\lambda_4 \le x \le \lambda_1 \\
&= [1-(x-\lambda_1)/(\lambda_2(1-\lambda_4))]^{(1-\lambda_3)/\lambda_3}/(\lambda_2\lambda_3) && \text{if } \lambda_1 \le x \le \lambda_1+\lambda_2(1-\lambda_4) \qquad (1) \\
&= 0 && \text{otherwise}
\end{aligned}
$$

where $-\infty < \lambda_1 < \infty$, $0 < \lambda_2 < \infty$, $0 < \lambda_3 < \infty$ and $0 \le \lambda_4 \le 1$.

This family has value due to its ability to model a variety of probabilistic phenomena by varying only the four parameters.

There is much literature on methods for modeling random variables using simple, yet versatile, methods. Schmeiser [11], in a survey of these methods, in the context of computer simulation, discusses systems of distributions (Pearson and Johnson), approximations to the inverse distribution function (expansion techniques, polynominal regression, and rectangular approximation), and four parameter distributions (beta, four parameter gamma, Burr, generalized lambda, and absolute lambda).

The distribution developed in this paper is a four parameter family. As such, it is appropriate to survey the literature of four parameter distributions. Stacy [14], Stacy and Mihram [15] and Harter [4] discuss the four parameter gamma distribution, which includes both the three parameter Weibull and gamma distribution. This distribution is useful for modelling life times with range $[0,\infty)$.

However, parameter estimation is not straight-forward. Burr [1,2] developed a more flexible distribution with the same range and straight-forward random variable generation. Disadvantages are that both heavy and light tailed distributions are not obtainable, nor are symmetric distributions. Ramberg and Schmeiser [7, 8] generalized Tukey's lambda distribution [18] to obtain a family for all but light tailed distributions, including symmetric distributions. The exponential distribution is a limiting case and the normal distribution function is approximated within a tolerance of .001. (See Joiner and Rosenblatt [5] for a related approximation.) Random variate generation is straight-forward. However, parameter estimation requires tables such as given in [3] and [10] for matching moments. Other estimation techniques, such as maximum likelihood estimation, are not straight-forward due to the distribution function and density function not being expressable in closed form.

Schmeiser and Deutsch [13] describe a family of distributions which can be used to obtain a distribution having any first four moments. The exponential distribution is a limiting case. Closed form expressions are given to calculate parameter values to match desired mode, percentile of the mode and any other two quantiles. Moments may be matched graphically. (No tables exist to date.) Random variable generation is closed form and requires only one exponential operation. The disadvantage of this distribution is that while any four independent properties may be specified, the shape of the distribution is most satisfactory as a quick and dirty technique. The shortcomings involve truncated tails and a density function value at

the mode which takes on only the values zero, one, and infinity.

The family of distributions developed in the following sections provides a model which allows for closed form parameter determination in some cases, has a finite density at the mode, does not truncate the tails and provides a model for any mean and variance and a wide range of commonly used third and fourth moments.

2. PROPERTIES OF THE FAMILY

This section developes the properties of the family having density function given in equation (1). From the density function, the distribution function is seen to be:

$$
\begin{aligned}
F(x) &= 0 && \text{if } x < \lambda_1 - \lambda_2 \lambda_4 \\
&= \lambda_4[1+(x-\lambda_1)/(\lambda_2\lambda_4)]^{1/\lambda_3} && \text{if } \lambda_1 - \lambda_2\lambda_4 \le x \le \lambda_1 \\
&= 1-(1-\lambda_4)[1-(x-\lambda_1)/(\lambda_2(1-\lambda_4))]^{1/\lambda_3} && \text{if } \lambda_1 < x \le \lambda_1 + \lambda_2(1-\lambda_4) \\
&= 1 && \text{if } x > \lambda_1 + \lambda_2(1-\lambda_4)
\end{aligned}
$$

From the density and distribution functions, the role of each of the four parameters is clear. The parameters λ_1 and λ_2 determine location and scaling, respectively. More specifically, the single mode or anti-mode is at λ_1 and λ_4 is the probability that an observed value is less than or equal to λ_1. Thus λ_4 may be used to determine the degree of asymmetry with $\lambda_4 = .5$ yielding symmetric distributions. Finally λ_3 corresponds most closely to the distribution tail weight. In combination, λ_3 and λ_4 determine the third and fourth standardized moments.

That these parameters yield a wide range of shapes can be seen by considering some special cases. Bernoulli trials and the Laplace (double exponential) distribution are symmetric limiting cases. As λ_3 approaches infinity with $\lambda_1 = \lambda_4$ and $\lambda_2 = 1$, the probability that $X = 0$ is λ_4 and the probability that $X = 1$ is $1 - \lambda_4$. As λ_3 approaches zero, the limiting distribution is the Laplace. Since Bernoulli trials correspond to the minimum possible fourth moment for a given third moment and the Laplace has a fourth standardized moment (kurtosis) of six, it is seen that a wide range of distribution shapes may be obtained.

Another important asymmetric limiting case is the exponential. By letting $\lambda_1 = 0$, $\lambda_2 = \lambda/\lambda_3$ and $\lambda_4 = 0$, the limiting case as λ_3 approaches zero is exponential with mean λ, as can be seen from

$$\lim_{\lambda_3 \to 0} F(x) = \lim_{\lambda_3 \to 0} 1 - [1 - x/(\lambda/\lambda_3)]^{1/\lambda_3} \qquad 0 \le x < \infty$$

$$= 1 - e^{-x/\lambda} \qquad 0 \le x < \infty$$

which is the distribution function of the exponential distribution having mean λ.

Additional special cases are the power series distribution having density function

$$f(x) = \lambda\, x^{\lambda-1} \qquad 0 \le x \le 1$$

$$= 0 \qquad \text{elsewhere}$$

obtained when $\lambda_1 = \lambda_2 = \lambda_4 = 1$ and $\lambda_3 = 1/\lambda$. The uniform distribution over the interval $[a,b]$ is obtained by setting $\lambda_1 = a + (b-a)\lambda_4$, $\lambda_2 = b-a$, $\lambda_3 = 1$ and λ_4 to any value in the unit interval [0,1]. Both the power series distribution and the uniform are special cases of the beta distribution.

Some graphs may illustrate the versatility of the family. Figure 1 shows graphs of the distribution for $\lambda_4 = .5$ for various values of λ_3. Changing λ_1 and/or λ_2 would change the location and/or scale, but not the shape, of the distribution. Since $\lambda_4 = .5$, the distribution is symmetric for all λ_3. Figure 2 shows graphs corresponding to $\lambda_4 = .75$. Here 75% of the area under each curve lies to the left of the mode (anti-mode) λ_1 for all λ_3.

FIGURES 1 & 2 ABOUT HERE

The generation of random variates from this family is straight forward via the inverse distribution function

$$
\begin{aligned}
x = F^{-1}(p) &= \lambda_1 - \lambda_2\lambda_4[1-(p/\lambda_4)^{\lambda_3}] && \text{if } p \le \lambda_4 \\
&= \lambda_1 + \lambda_2(1-\lambda_4)[1-((1-p)/(1-\lambda_4))^{\lambda_3}] && \text{if } p > \lambda_4
\end{aligned}
\tag{2}
$$

Insertion of a U(0,1) (pseudo) random variate into the inverse distribution function generates a value x from the distribution

determined by λ_1, λ_2, λ_3, and λ_4. This generation is relatively fast in that it requires only one exponentiation. In addition, the closed-form inverse distribution function allows the direct generation of order statistics via the methods of Lurie and Hartley [6], Schmeiser [12], Ramberg and Tadikamalla [9] and Schucany [14].

The moments of the distribution are now derived. Letting $\lambda_1 = 0$,

$$E\,\{X^k | \lambda_1 = 0\} = \int_{-\lambda_4\lambda_2}^{(1-\lambda_4)\lambda_2} xf(x)\,dx$$

$$= \int_0^{\lambda_4} \{\lambda_2\lambda_4[(p/\lambda_4)^{\lambda_3} - 1]\}^k dp + \int_{\lambda_4}^1 \{\lambda_2(1-\lambda_4)[1-((1-p)/(1-\lambda_4))^{\lambda_3}]\}^k dp$$

$$= \lambda_2^k \{\lambda_4^k \int_0^{\lambda_4} [\sum_{i=0}^k (-1)^i \binom{k}{i} (p/\lambda_4)^{\lambda_3(k-i)}]dp$$

$$+ (1-\lambda_4)^k \int_{\lambda_4}^1 [\sum_{i=0}^k (-1)^i \binom{k}{i} ((1-p)/(1-\lambda_4))^{\lambda_3 i}]dp\}$$

$$= \lambda_2^k \{\lambda_4^k \sum_{i=0}^k (-1)^i \binom{k}{i} (\lambda_4)^{-\lambda_3(k-i)} \int_0^{\lambda_4} p^{\lambda_3(k-i)}\,dp$$

$$+ (1-\lambda_4)^k \sum_{i=0}^k (-1)^i \binom{k}{i} (1-\lambda_4)^{-\lambda_3 i} \int_{\lambda_4}^1 (1-p)^{\lambda_3 i}\,dp\}$$

$$= \lambda_2^k \{\lambda_4^{k+1} \sum_{i=0}^k \frac{(-1)^i \binom{k}{i}}{\lambda_3(k-i)+1} + (1-\lambda_4)^{k+1} \sum_{i=0}^k \frac{(-1)^i \binom{k}{i}}{\lambda_3\, i+1}\}$$

$$= \lambda_2^k \sum_{i=0}^{k} \frac{(-1)^i \binom{k}{i}}{\lambda_3 i + 1} \{(1-\lambda_4)^{k+1} + (-1)^k \lambda_4^{\,k+1}\}$$

$$= (\lambda_2\lambda_3)^k (k!) \; [\; \sum_{i=0}^{k} \binom{k+1}{i} (-\lambda_4)^i \,] \quad / \quad \prod_{i=1}^{k} (\lambda_3 i+1)$$

For any value of λ_1, some algebraic reduction yields

$$E\{X\} = \lambda_1 + \; [\lambda_2\lambda_3(1-2\lambda_4)]/(\lambda_3+1)$$

and

$$V\{X\} = (\lambda_2\lambda_3)^2 \quad [2\lambda_4(1-\lambda_4)(\lambda_3-1) + 1]/[(2\lambda_3 + 1)(\lambda_3 + 1)^2]$$

3. PARAMETER DETERMINATION

The four parameters of this family allow four independent properties to be specified. These properties may be moments, fractiles, upper and lower bounds, or mode when these properties are known. Alternately, the parameters may be estimated from sample data using the classical techniques of maximum likelihood, least squares, or method of moments, although these methods do not lead to closed-form solutions.

3.1 Obtaining Specified Properties When No Data is Available.

Often a distribution is desired which possesses certain properties. This happens when the only available data is from expert opinion as to, say, the mode, the fractile of the mode and bounds on the range of the variable. "No data" situations commonly occur in PERT/CPM

modeling where the activity has been performed seldom, if ever, before. Whenever a new system is being designed, and therefore no data is available, the use of specified properties to determine parameter values is necessary. For the family discussed in this paper, four independent properties may be specified, whereas for most commonly used distributions only one or two properties can be obtained.

Parameters for this family may be determined in a closed-form manner from the interpretation of the four parameters. In particular, letting m denote the mode, a the lower bound and b the upper bound,

$$\lambda_1 = m \tag{3}$$

$$\lambda_4 = \text{fractile of the mode } \lambda_1$$

$$\lambda_2 = b\text{-}a \tag{4}$$

$$\lambda_3 = \ln[1 + (x_1-\lambda_1)/\lambda_2\lambda_4]/\ln[p_1/\lambda_4] \qquad \text{if } p_1 \leq \lambda_4$$

$$= \ln[1-(x_1-\lambda_1)/(\lambda_2(1-\lambda_4))]/\ln[(1-p_1)/(1-\lambda_4)] \qquad \text{if } p_1 > \lambda_4$$

yields the fractile p_1 at $x = x_1$.

Other properties may be combined to yield parameter values. In particular

$$\lambda_1 - \lambda_2\lambda_4 = a, \tag{6}$$

$$\lambda_1 + \lambda_2(1-\lambda_4) = b,$$

$$\lambda_1 + \lambda_2\lambda_3(1-2\lambda_4)/(\lambda_3 + 1) = \text{mean}$$

and

$$(\lambda_2\lambda_3)^2\ [2\lambda_4(1-\lambda_4)(\lambda_3 -1) + 1]/[(2\lambda_3 + 1)(\lambda_3 + 1)^2] = \text{variance}$$

3.2 Parameter Estimation from Data.

When data $x_1, x_2, \ldots, x_n$ are available, the maximum likelihood estimates may be used. Differentiating the likelihood function with respect to each of the four parameters leads to

$$\lambda_1: \quad \sum_{i=0}^{n_1} [\lambda_4 + z_i]^{-1} = \sum_{i=n_1+1}^{n} [1-\lambda_4 - z_i]^{-1}$$

$$\lambda_2: \quad n\,\lambda_3/(\lambda_3 - 1) = \sum_{i=1}^{n_1} z_i/(\lambda_4 + z_i) - \sum_{i=n_1+1}^{n} z_i/(1-\lambda_4 - z_i)$$

$$\lambda_3: \quad \lambda_3 = -\{\sum_{i=1}^{n_1} \ln[1+z_i/\lambda_4] + \sum_{i=n_1+1}^{n} \ln[1-z_i/(1-\lambda_4)]\}/n \qquad (7)$$

$$\lambda_4: \quad (\lambda_4-1)/\lambda_4 = \{\sum_{i=n_1+1}^{n} z_i/(1-\lambda_4-z_i)\}/\sum_{i=1}^{n_1} z_i/(\lambda_4+z_i) \qquad (8)$$

where $z_i = (x_i - \lambda_1)/\lambda_2$ and n_1 is the number of x_i values less than λ_1.

Some care must be taken in using these maximum likelihood equations, as with any family of distributions which includes points at which the density function is infinite. Taken literally, the equations always will lead to a U-shaped distribution ($\lambda_3 > 1$) with end points at $x_{(1)}$ and $x_{(n)}$ (the first and nth order statistics), since the likelihood function is then infinite. When the distribution is

known not to be U shaped, the estimates can be improved by specifying the lower bound a and the upper bound b. Then λ_2 = b-a and $\lambda_1 = a + (b-a)\lambda_4$.

A modified maximum likelihood algorithm is then

1. Set λ_2 = b-a, from equation (4).
2. Solve for λ_4, in equation (8), keeping $\lambda_1 = a + (b-a)\lambda_4$ from equation (6) and $n_1 = [n\lambda_4]$. (Any standard unidimensional search technique, such as binary search, may be used.)
3. Set $\lambda_3 = -\{\sum_{i=1}^{n_1} \ln [1+z_i/\lambda_4] + \sum_{i=n_1+1}^{n} \ln[1-z_i/(1-\lambda_4)]\}/n$

4. EXAMPLE

In modeling a logic network, the propogation delay of each component is of interest. Using the Texas Instruments TTL Data Book [17], the minimum time, typical time and maximum time for a positive-NAND gate is 2, 7 and 15 nannoseconds (ns), respectively. In a particular situation, it is desired to have 1% of the gates exceeding the stated maximum of 15 ns with the true maximum being 20 ns. This can be modeled using parameter values

λ_1 = 7 ns from equation (3),

λ_2 = range = 18 ns from equation (4),

λ_4 = .278 from equation (6),

and

$$\lambda_3 = .2234 \qquad \text{from equation (5).}$$

If simulation is to be performed, random variates may be generated by substituting U(0,1) values p into equation (2)

$$x = 7 - 18\,(.278)\,[1 - (p/.278)^{.2234}] \qquad \text{if } p \leq .278$$

$$= 7 + 18\,(.722)\,[1 - ((1 - p)/.722)^{.2234}] \qquad \text{if } p > .278.$$

5. SUMMARY

A four parameter family has been developed which has several appealing properties when used for probabilistic modeling. In addition to the versatility in terms of the shapes which the distributions assume, the density function, cdf, inverse cdf, and moments are all closed form. Parameter determination is closed form in many situations where specified properties, such as mode and fractiles, are required. Maximum likelihood estimation is considered and an example is given.

6. ACKNOWLEDGEMENTS

This research was supported by Office of Naval Research contract N00014-C-77-0425, Code 431.

REFERENCES

[1]. BURR, I. W. (1942). Cumulative frequency functions, Annals of Mathematical Statistics, 13, 215-232.

[2]. BURR, I. W. (1973). Parameters for a general system of distributions to match a grid of α_3 and α_4. Communications in Statistics, 2, 1-21.

[3]. DUDEWICZ, E. J., RAMBERG, J. S., and TADIKAMALLA, P. R. (1974). A distribution for data fitting and simulation. ASQC Transactions, 28, 407-418.

[4]. HARTER, H. L. (1967). Maximum likelihood estimation of the parameters of a four-parameter gamma population from complete and censored samples. Technometrics, 9, 159-165.

[5]. JOINER, B. L., and ROSENBLATT, J. R. (1971). Some properties of the range in samples from Tukey's symmetric lambda distribution. Journal of the American Statistical Association, 66, 398-399.

[6]. LURIE, D. and HARTLEY, H. O. (1972). Machine-generation of order statistics for Monte Carlo computations, The American Statistician, 26, 26-27.

[7]. RAMBERG, J. S. and SCHMEISER, B. W., (1972). An approximate method for generating asymmetric random variables. Communications of the ACM. 15, 987-990.

[8]. RAMBERG, J. S. and SCHMEISER, B. W., (1974). An approximate method for generating asymmetric random variables. Communications of the ACM, 17, 78-82.

[9]. RAMBERG, J. S. and TADIKAMALLA, P. R. (1978). On the generation of subsets of order statistics. Journal of Statistical Computation and Simulation, 6, 239-241.

[10]. RAMBERG, J. S., TADIKAMALLA, P. R., DUDEWICZ, E. J., and MYKTKA, E. F., (1978). A probability distribution for fitting empirical data which simplifies applications. Technometrics, to appear.

[11]. SCHMEISER, B. W. (1977). Methods for modelling and generating probabilistic components in digital computer simulation when the standard distributions are not adequate: A Survey. Proceedings of the Winter Simulation Conference, 51-55.

[12]. SCHMEISER, B. W., (1978). Generation of the maximum (minimum) value in digital computer simulation. Journal of Statistical Computation and Simulation, 8, 103-116.

[13]. SCHMEISER, B. W. and DEUTSCH, S. J. (1977). A versatile family of of probability distributions suitable for simulation. AIIE Transactions, 9, 176-182.

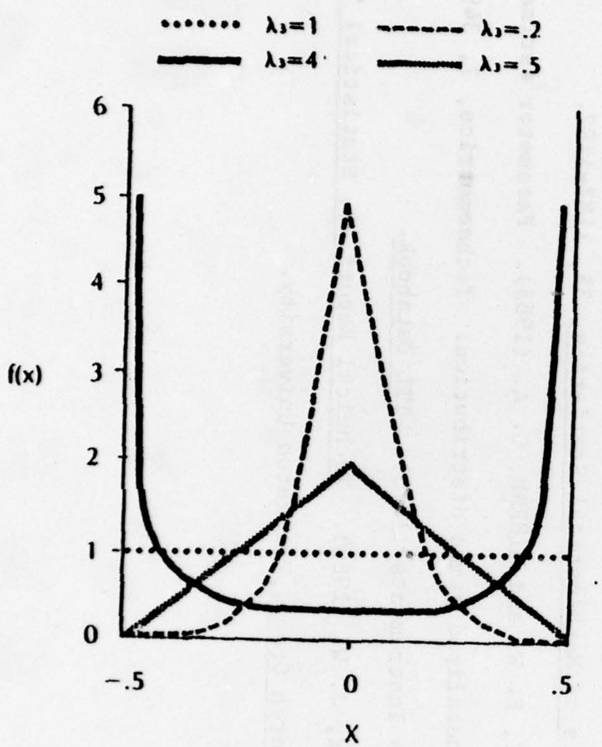
[14]. SCHUCANY, W. R. (1972). Order statistics in simulation. Journal of Statistical Computation and Simulation, 1, 281-286.

[15]. STACY, E. W. (1962). A generalization of the gamma distribution. Annals of Mathematical Statistics, 33, 1187-1192.

[16]. STACY, E. W. and MIHRAM, G. A. (1965). Parameter estimation for a generalized gamma distribution. Technometrics, 7, 349-358.

[17]. Texas Instruments (1973). TTL Databook.

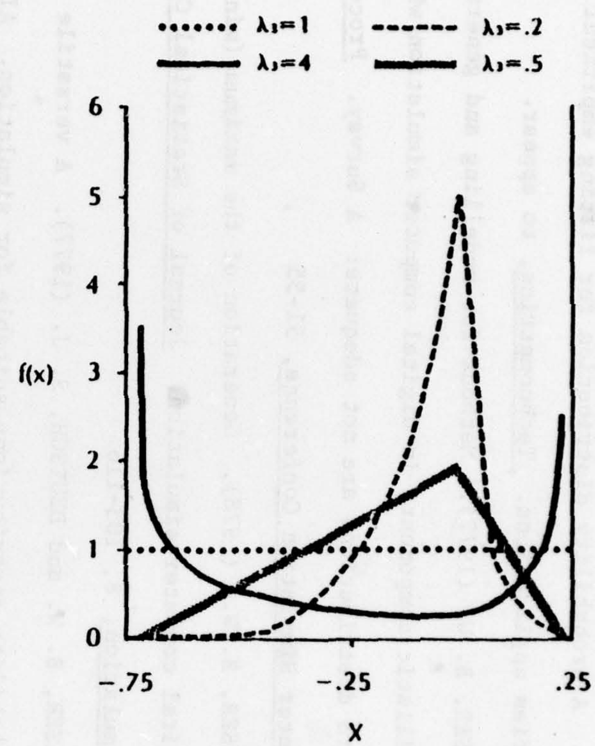
[18]. TUKEY, J. W. (1960). Technical Report 36, Statistical Techniques Research Group, Princeton University.



SYMMETRIC EXAMPLES

$\lambda_1=0 \quad \lambda_2=1 \quad \lambda_4=.5$

FIGURE 1.



ASYMMETRIC EXAMPLES

$\lambda_1=0 \quad \lambda_2=1 \quad \lambda_4=.75$

FIGURE 2.

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER 14 OREM-78015	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
4. TITLE ~~(and Subtitle)~~ 6 Another Versatile Family of Probability Distributions		5. TYPE OF REPORT & PERIOD COVERED 9 Technical Report ~~6. PERFORMING ORG. REPORT~~ NUMBER
7. AUTHOR(s) 10 Bruce W. Schmeiser Ram Lal		8. CONTRACT OR GRANT NUMBER(s) 15 N00014-77-C-0425
9. PERFORMING ORGANIZATION NAME AND ADDRESS Southern Methodist University Dallas, TX 75275		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS NR 277-236
11. CONTROLLING OFFICE NAME AND ADDRESS Naval Analysis Program - Code 431 Office of Naval Research Arlington, VA 22217		12. ~~REPORT DATE~~ 11 Dec~~ember~~ 1978 13. ~~NUMBER OF PAGES~~ 16
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office) 12 17 p'		15. SECURITY CLASS. (of this report) Unclassified 15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Family of Distributions, Simulation, Process generation, Modeling

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

A new four parameter family of probability distributions is described. Special cases include Bernoulli trials, uniform, power series, exponential, triangular and Laplace (double exponential) distributions. Statistical properties, parameter determination, and random variate generation are developed.

410 597

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)